UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P O Box 1450
Alexandria, Virginia 22313-1450
www uspto gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/578,640	05/09/2006	Sven C. Martin	DE 030375	1743

24737 7590 03/26/2010
PHILIPS INTELLECTUAL PROPERTY & STANDARDS
P.O. BOX 3001
BRIARCLIFF MANOR, NY 10510

EXAMINER	
COLUCCI, MICHAEL C	

ART UNIT	PAPER NUMBER
2626	

MAIL DATE	DELIVERY MODE
03/26/2010	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/578,640	MARTIN, SVEN C.	
	Examiner	Art Unit	
	MICHAEL C. COLUCCI	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *02/05/2010*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1 and 4-18* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1 and 4-18* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)

2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.

4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .

5) ☐ Notice of Informal Patent Application

6) ☐ Other: _____.

DETAILED ACTION

Response to Arguments

1. Applicant's arguments filed 02/05/2010 have been fully considered but they are not persuasive. In response to the amendments filed 02/05/2010 in claims 1, 6, and 11, Examiner has maintained the use of Brill in view of Schabes and Papineni. Examiner has also used these same combination of references for the rejection of new claims 16-18. Examiner believes that a weak annotation and an inclusion/exclusion list are not necessarily well known in the art and therefore looks to the disclosure. Examiner concurs that there is support given in the specification within Remarks filed 02/05/2010 referencing [0015], which states:

"*The probabilistic dependency between phrases and tags is further denoted as mapping probability and its determination is based on the training corpus of sentences. Initially, the method has no information about the annotation between tags and phrases of the training corpus. <u>In order to perform a calculation of the mapping probability a weak annotation between phrases and semantic tags must be somehow provided</u>. <u>Such a weak annotation can be realized for example by assigning a set of candidate semantic tags to a phrase</u>. Alternatively an IEL (inclusion/exclusion list) can be used. An IEL represents a list that <u>includes or excludes</u> various semantic tags that can be mapped or must not map a phrase*". (Present invention spec. [0015]).

Examiner does not believe a "weak annotation" and "(inclusion/exclusion list)" to be well known in the art, therefore based on the cited support in the disclosure of the

present invention, Examiner understands the concept of a weak annotation to be that which provides *a set of candidate semantic tags to a phrase* which allows for the calculation of a mapping probability. Further, Examiner understands an IEL to be a list that *includes or excludes various semantic tags that can be mapped or must not map a phrase*. Examiner believes that Schabes improves the probabilistic natural language understanding via an expectation-maximization algorithm of Brill to include words within a sentence that are annotated with a tag, such as a part-of-speech tag (Schabes Col. 23 lines 18-25).

Schabes also teaches a listing of words that DO and DO NOT correspond to a proper relationship, wherein Schabes teaches the break down of the proper output of tags, where order to display the entries from the dictionary that correspond to the context, all the entries in a dictionary 970 that correspond to a root found in the set 37 of pairs of roots and parts-of-speech that correspond to the context 950 are displayed at 980. In the above example, all entries for the verb "leave" will be displayed as entries relevant to the context. In order to display the entries from the dictionary that do not correspond to the context, all the entries in the dictionary 970 that correspond to a root found in the set of pairs of roots and parts-of-speech that do not correspond to the context 960 are displayed at 990. In the above example, all entries for the word "left" as an adjective, as an adverb and as a singular noun are displayed as entries not relevant to the context such as elements 980 and 990 (SchabesCol. 26 lines 12-21 & Fig. 14).

Initially, Brill teaches an expectation-maximization algorithm, wherein those skilled in the art will recognize that the present invention can be applied to any of the

trainable natural language components that are present in a natural language understanding unit. Under the method of the present invention, one or more of the specifications 324, 326, 334 and/or 336 are adjusted through unsupervised training. In the description below, <u>an unsupervised training method involving generating and testing candidate learning sets is described</u>. However, those skilled in the art will recognize that <u>the present invention may be incorporated in other unsupervised training techniques such as greedy hill climbing and variants of the expectation-maximization algorithm</u> (Brill [0027-0028]).

Examiner believes that Schabes improves the candidate sets and expectation-maximization algorithm of Brill to allow for a set of sentences in which the words of each sentence are annotated with their part-of-speech tags (Schabes Col. 23 lines 18-25) that improve a NLU expectation-maximization algorithm to retain or omit sets of tags (SchabesCol. 26 lines 12-21).

Claim Rejections - 35 USC § 103

2. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

3. Claims 1 and 4-18 are rejected under 35 U.S.C. 103(a) as being unpatentable over Brill et al. US 20020169596 A1 (hereinafter Brill) in view of Schabes et al. US 5537317 A (hereinafter Schabes) and further in view of Papineni et al. US 5991710 A (hereinafter Papineni).

Re claim 1, Brill teaches a method carried out by a processor, comprising:

extracting a phrase from a training corpus ([0021], semantic interpreter analyzing sentences from a corpus);

calculating a probability that the phrase is mapped to a semantic tag ([0025], semantic interpreter mapping components) from a list of unordered semantic tags;

mapping the phrase to the semantic tag ([0033-0034], highest score for learning set) with the highest mapping probability ([0028] maximization algorithm);

generating a mapping table containing the phrase and its corresponding semantic tag ([0025], semantic interpreter mapping components)

However Brill fails to teach calculating a probability that the phrase is mapped to a semantic tag from a list

wherein a weak annotation between the phrase and the semantic tag is provided to the processor

Schabes teaches past limitations and an improvement upon them, wherein Schabes teaches that in the past, in order to ascertain proper usage, the grammaticality of a sentence was computed as the probability of this sentence to occur in English. Such statistical approach assigns high probability to grammatically correct sentences,

and low probability to ungrammatical sentences. The statistical is obtained by training on a collection of English sentences, or a training corpus. The corpus defines correct usage. As a result, when a sentence is typed in to such a grammar checking system, the probability of the entire sentence correlating with the corpus is computed. It will be appreciated in order to entertain the entire English vocabulary, about 60,000 words, a corpus of at several hundred trillion words must be used. Furthermore, a comparable number of probabilities must be stored on the computer. Thus the task of analyzing entire sentences is both computationally and storage intensive. In order to establish correct usage in the Subject System, it is the probability of a sequence of parts of speech which is derived. For this purpose, one can consider that there are between 100 and 400 possible parts of speech depending how sophisticated the system is to be. This translates to a several million word training corpus as opposed to several hundred trillion. This type of analysis can be easily performed on standard computing platforms including the ones used for word processing. Thus in the subject system, a sentence is first broken up into parts of speech. For instance, the sentence "I heard this band play" is analyzed as follows: PRONOUN, VERB, DETERMINER, NOUN, VERB. The probability of this part of speech sequence, is determined by comparing the sequence to the corpus. This is also not feasible unless one merely consider the so-called tri-grams. Tri-grams are triple of parts of speech which are adjacent in the input sentence. Analyzing three adjacent parts of speech is usually sufficient to establish correctness; and it the probability of these tri-grams which is utilized to establish that a particular sentence involves correct usage. Thus rather than checking the entire sentence, the

probability of three adjacent parts of speech is computed from the training corpus (Schabes Col. 8 lines 13-51).

Further, Schabes teaches that the entries of a dictionary are selected and ranked based on the part of speech assigned to the given word in context. The entries corresponding to the word in context are first selected. The other entries not relevant to the current context are still available at the request of the user. The part of speech of the given word in context is disambiguated with the part of speech tagger described above. By way of illustration, assuming the word "left" in the sentence "He left a minute ago", the part of speech tagger assigns the tag "verb past tense" for the word "left" in that sentence. For this case, the Subject System selects the entries for the verb "leave" corresponding to the usage of "left" in that context and then selects the entries for "left" not used in that context, in particular the ones for "left" as an adjective, as an adverb and as a noun (Schabes Col. 24 lines 45-60).

Schabes teaches a sentence that is annotated with a tag, such as a part-of-speech tag (Schabes Col. 23 lines 18-25).

Schabes also teaches a listing of words that DO and DO NOT correspond to a proper relationship, wherein Schabes teaches the break down of the proper output of tags, where order to display the entries from the dictionary that correspond to the context, all the entries in a dictionary 970 that correspond to a root found in the set 37 of pairs of roots and parts-of-speech that correspond to the context 950 are displayed at 980. In the above example, all entries for the verb "leave" will be displayed as entries relevant to the context. In order to display the entries from the dictionary that do not

correspond to the context, all the entries in the dictionary 970 that correspond to a root found in the set of pairs of roots and parts-of-speech that do not correspond to the context 960 are displayed at 990. In the above example, all entries for the word "left" as an adjective, as an adverb and as a singular noun are displayed as entries not relevant to the context (SchabesCol. 26 lines 12-21).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Brill to incorporate calculating a probability that the phrase is mapped to a semantic tag from a list of semantic tags wherein a weak annotation between the phrase and the semantic tag is provided to the processor as taught by Schabes to allow for the tagging of semantic portions of a sentence (such as parts of speech) in order to prioritize (i.e. the best ranking/probability) semantic tags within a sentence to maintain the proper context based on adjacent tags in a sentence (Schabes Col. 24 lines 45-60) and to further allow for a set of sentences in which the words of each sentence are annotated with their part-of-speech tags (Schabes Col. 23 lines 18-25) that improve a NLU expectation-maximization algorithm to retain or omit sets of tags such as elements 980 and 990 (SchabesCol. 26 lines 12-21 & Fig. 14).

However, Brill in view of Schabes fails to teach the use of semantic unordered lists.

Papineni teaches the identification of word mapping relative to an unordered list of grammatical components, wherein word-set feature functions formed and supported by the translation model of the present invention are characterized such that s and t are

unordered sets of words. That is, s is in S if all n words of s are in S, regardless of the order in which they occur in S. Likewise, t is in T if all n words of t are in T, regardless of the order in which they occur in T. An example of a word-set feature function or operation performed by the model in the ATIS domain would be searching for the existence of the unordered words "departing" and "after" among the formal sentence candidates (stored in target language candidate store 30), given an English sentence having the unordered words "leave" and "after" contained therein. For instance given the sample English sentences (E.sub.1 through E.sub.6) and the sample formal sentences (F.sub.1 through F.sub.5) above, the word-set feature function fires on E.sub.1 and F.sub.1, thus, identifying the pair (E.sub.1, F.sub.1). The same is true for the pair (E.sub.2, F.sub.1) (Col. 5 lines 45 – Col. 6 line 50).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Brill in view of Schabes to incorporate calculating a probability that the phrase is mapped to a semantic tag from a list of unordered semantic tags as taught by Papineni to allow for the identification of all words found within a set of words regardless of order/sequence of words in a phrase or group of words Col. 5 lines 45 – Col. 6 line 50).

Re claims 6, and 11, Brill teaches a processor executing a computer program product to:

calculate a mapping probability that a semantic tag of a set of candidate semantic tags is assigned to a phrase ([0025]), wherein the calculation of the mapping

probability is performed by means of a statistical procedure based on a set of phrases constituting a corpus of sentences ([0024]), each of the phrases having assigned a set of candidate semantic tags ([0028]).

generate a mapping table from the performed mapping ([0035])

However, Brill fails to teach mapping probability that is performed by means of a statistical procedure based on a set of phrases

However Brill fails to teach calculating a probability from a list of semantic tags

Schabes teaches past limitations and an improvement upon them, wherein Schabes teaches that in the past, in order to ascertain proper usage, the grammaticality of a sentence was computed as the probability of this sentence to occur in English. Such statistical approach assigns high probability to grammatically correct sentences, and low probability to ungrammatical sentences. The statistical is obtained by training on a collection of English sentences, or a training corpus. The corpus defines correct usage. As a result, when a sentence is typed in to such a grammar checking system, the probability of the entire sentence correlating with the corpus is computed. It will be appreciated in order to entertain the entire English vocabulary, about 60,000 words, a corpus of at several hundred trillion words must be used. Furthermore, a comparable number of probabilities must be stored on the computer. Thus the task of analyzing entire sentences is both computationally and storage intensive. In order to establish correct usage in the Subject System, it is the probability of a sequence of parts of speech which is derived. For this purpose, one can consider that there are between 100 and 400 possible parts of speech depending how sophisticated the system is to be.

This translates to a several million word training corpus as opposed to several hundred trillion. This type of analysis can be easily performed on standard computing platforms including the ones used for word processing. Thus in the subject system, a sentence is first broken up into parts of speech. For instance, the sentence "I heard this band play" is analyzed as follows: PRONOUN, VERB, DETERMINER, NOUN, VERB. The probability of this part of speech sequence, is determined by comparing the sequence to the corpus. This is also not feasible unless one merely consider the so-called tri-grams. Tri-grams are triple of parts of speech which are adjacent in the input sentence. Analyzing three adjacent parts of speech is usually sufficient to establish correctness; and it the probability of these tri-grams which is utilized to establish that a particular sentence involves correct usage. Thus rather than checking the entire sentence, the probability of three adjacent parts of speech is computed from the training corpus (Schabes Col. 8 lines 13-51).

Further, Schabes teaches that the entries of a dictionary are selected and ranked based on the part of speech assigned to the given word in context. The entries corresponding to the word in context are first selected. The other entries not relevant to the current context are still available at the request of the user. The part of speech of the given word in context is disambiguated with the part of speech tagger described above. By way of illustration, assuming the word "left" in the sentence "He left a minute ago", the part of speech tagger assigns the tag "verb past tense" for the word "left" in that sentence. For this case, the Subject System selects the entries for the verb "leave" corresponding to the usage of "left" in that context and then selects the entries for "left"

not used in that context, in particular the ones for "left" as an adjective, as an adverb and as a noun (Schabes Col. 24 lines 45-60).

Schabes also teaches well known previous techniques, wherein in the past, in order to ascertain proper usage, the grammaticality of a sentence was computed as the probability of this sentence to occur in English. Such statistical approach assigns high probability to grammatically correct sentences, and low probability to ungrammatical sentences. The statistical is obtained by training on a collection of English sentences, or a training corpus. The corpus defines correct usage. As a result, when a sentence is typed in to such a grammar checking system, the probability of the entire sentence correlating with the corpus is computed. It will be appreciated in order to entertain the entire English vocabulary, about 60,000 words, a corpus of at several hundred trillion words must be used. Furthermore, a comparable number of probabilities must be stored on the computer. Thus the task of analyzing entire sentences is both computationally and storage intensive (Schabes Col. 8 lines 12-28).

Further, Schabes overcomes previous techniques, wherein rather than comparing the above mentioned probabilities, in a preferred embodiment, the subject system compares the geometric average of these probabilities by taking into account their word lengths, i.e. by comparing the logarithm of P1 divided by the number of words in S1, and the logarithm of P2 divided by the number of words in S2. This is important in cases where a single word may be confused with a sequence of words such as "maybe" and "may be". Directly comparing the probabilities of the part of speech sequences would favor shorter sentences instead of longer sentences, an not

necessarily correct result, since the statistical language model assigns lower probabilities to longer sentences (Schabes Col. 9 lines 55-67).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Brill to incorporate mapping probability that is performed by means of a statistical procedure based on a set of phrases and semantic tags assigned to a phrase as taught by Schabes to allow for the recognition of parts of speech and individual in addition to the identification of sentences/phrases, wherein higher/lower probabilities are assigned to sentences and the length of the sentences in an unsupervised or even supervised system (Schabes Col. 9 lines 55-67) and to further allow for the tagging of semantic portions of a sentence (such as parts of speech) in order to prioritize (i.e. the best ranking/probability) semantic tags within a sentence to maintain the proper context based on adjacent tags in a sentence (Schabes Col. 24 lines 45-60).

However, Brill in view of Schabes fails to teach the use of semantic unordered lists.

Papineni teaches the identification of word mapping relative to an unordered list of grammatical components, wherein word-set feature functions formed and supported by the translation model of the present invention are characterized such that s and t are unordered sets of words. That is, s is in S if all n words of s are in S, regardless of the order in which they occur in S. Likewise, t is in T if all n words of t are in T, regardless of the order in which they occur in T. An example of a word-set feature function or

operation performed by the model in the ATIS domain would be searching for the existence of the unordered words "departing" and "after" among the formal sentence candidates (stored in target language candidate store 30), given an English sentence having the unordered words "leave" and "after" contained therein. For instance given the sample English sentences (E.sub.1 through E.sub.6) and the sample formal sentences (F.sub.1 through F.sub.5) above, the word-set feature function fires on E.sub.1 and F.sub.1, thus, identifying the pair (E.sub.1, F.sub.1). The same is true for the pair (E.sub.2, F.sub.1) (Col. 5 lines 45 – Col. 6 line 50).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Brill in view of Schabes to incorporate calculating a probability that the phrase is mapped to a semantic tag from a list of unordered semantic tags as taught by Papineni to allow for the identification of all words found within a set of words regardless of order/sequence of words in a phrase or group of words Col. 5 lines 45 – Col. 6 line 50).

Re claims 7 and 12, Brill teaches the method according to claim I, for each phrase further comprising calculating a set of mapping probabilities ([0025]), providing the probability for each semantic tag of the set of candidate semantic tags being assigned to the phrase ([0028]).

However, Brill fails to teach providing the probability for each semantic tag of the set of candidate semantic tags

Schabes teaches well known previous techniques, wherein in the past, in order to ascertain proper usage, the grammaticality of a sentence was computed as the probability of this sentence to occur in English. Such statistical approach assigns high probability to grammatically correct sentences, and low probability to ungrammatical sentences. The statistical is obtained by training on a collection of English sentences, or a training corpus. The corpus defines correct usage. As a result, when a sentence is typed in to such a grammar checking system, the probability of the entire sentence correlating with the corpus is computed. It will be appreciated in order to entertain the entire English vocabulary, about 60,000 words, a corpus of at several hundred trillion words must be used. Furthermore, a comparable number of probabilities must be stored on the computer. Thus the task of analyzing entire sentences is both computationally and storage intensive (Schabes Col. 8 lines 12-28).

Further, Schabes overcomes previous techniques, wherein rather than comparing the above mentioned probabilities, in a preferred embodiment, the subject system compares the geometric average of these probabilities by taking into account their word lengths, i.e. by comparing the logarithm of P1 divided by the number of words in S1, and the logarithm of P2 divided by the number of words in S2. This is important in cases where a single word may be confused with a sequence of words such as "maybe" and "may be". Directly comparing the probabilities of the part of speech sequences would favor shorter sentences instead of longer sentences, an not necessarily correct result, since the statistical language model assigns lower probabilities to longer sentences (Schabes Col. 9 lines 55-67).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Brill to incorporate the probability for each semantic tag of the set of candidate semantic tags as taught by Schabes to allow for the recognition of parts of speech and individual in addition to the identification of sentences/phrases, wherein higher/lower probabilities are assigned to sentences and the length of the sentences in an unsupervised or even supervised system (Schabes Col. 9 lines 55-67).

Re claims 8 and 13, Brill teaches the method according to claim 2, further comprising determining one semantic tag of the set of candidate semantic tags ([0025]) having the highest mapping probability of the set of mapping probabilities and mapping the one semantic tag to the phrase ([0024])

However, Brill fails to teach determining one semantic tag of the set of candidate semantic tags having the highest mapping probability

Schabes teaches well known previous techniques, wherein in the past, in order to ascertain proper usage, the grammaticality of a sentence was computed as the probability of this sentence to occur in English. Such statistical approach assigns high probability to grammatically correct sentences, and low probability to ungrammatical sentences. The statistical is obtained by training on a collection of English sentences, or a training corpus. The corpus defines correct usage. As a result, when a sentence is typed in to such a grammar checking system, the probability of the entire sentence correlating with the corpus is computed. It will be appreciated in order to entertain the

entire English vocabulary, about 60,000 words, a corpus of at several hundred trillion words must be used. Furthermore, a comparable number of probabilities must be stored on the computer. Thus the task of analyzing entire sentences is both computationally and storage intensive (Schabes Col. 8 lines 12-28).

Further, Schabes overcomes previous techniques, wherein rather than comparing the above mentioned probabilities, in a preferred embodiment, the subject system compares the geometric average of these probabilities by taking into account their word lengths, i.e. by comparing the logarithm of P1 divided by the number of words in S1, and the logarithm of P2 divided by the number of words in S2. This is important in cases where a single word may be confused with a sequence of words such as "maybe" and "may be". Directly comparing the probabilities of the part of speech sequences would favor shorter sentences instead of longer sentences, an not necessarily correct result, since the statistical language model assigns lower probabilities to longer sentences (Schabes Col. 9 lines 55-67).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Brill to incorporate the probability for each semantic tag of the set of candidate semantic tags as taught by Schabes to allow for the recognition of parts of speech and individual in addition to the identification of sentences/phrases, wherein higher/lower probabilities are assigned to sentences and the length of the sentences in an unsupervised or even supervised system (Schabes Col. 9 lines 55-67).

Re claims 4, 9, and 14, Brill teaches the method according to claim 1, wherein the statistical procedure comprises an expectation maximization algorithm ([0028]).

Re claims 5, 10, and 15, Brill teaches the method according to claim 3 or 4, further comprising storing of performed mappings between a candidate semantic tag ([0025]) and a phrase in form of a mapping table ([0024]) in order to derive a grammar being applicable to unknown sentences or unknown phrases.

However, Brill fails to teach deriving a grammar being applicable to unknown sentences or unknown phrases

Schabes teaches well known previous techniques, wherein in the past, in order to ascertain proper usage, the grammaticality of a sentence was computed as the probability of this sentence to occur in English. Such statistical approach assigns high probability to grammatically correct sentences, and low probability to ungrammatical sentences. The statistical is obtained by training on a collection of English sentences, or a training corpus. The corpus defines correct usage. As a result, when a sentence is typed in to such a grammar checking system, the probability of the entire sentence correlating with the corpus is computed. It will be appreciated in order to entertain the entire English vocabulary, about 60,000 words, a corpus of at several hundred trillion words must be used. Furthermore, a comparable number of probabilities must be stored on the computer. Thus the task of analyzing entire sentences is both computationally and storage intensive (Schabes Col. 8 lines 12-28).

Further, Schabes overcomes previous techniques, wherein rather than comparing the above mentioned probabilities, in a preferred embodiment, the subject system compares the geometric average of these probabilities by taking into account their word lengths, i.e. by comparing the logarithm of P1 divided by the number of words in S1, and the logarithm of P2 divided by the number of words in S2. This is important in cases where a single word may be confused with a sequence of words such as "maybe" and "may be". Directly comparing the probabilities of the part of speech sequences would favor shorter sentences instead of longer sentences, an not necessarily correct result, since the statistical language model assigns lower probabilities to longer sentences (Schabes Col. 9 lines 55-67).

Furthermore, Schabes teaches that in particular importance in grammar checking is the ability to detect the sequence of parts of speech as they exist in a given sentence. Correct sentences will have parts of speech which follow a normal sequence, such that by analyzing the parts of speech sequence one can detect the probability that the sentence is correct in terms of its grammar. While prior art systems have tagged a sentence for parts of speech and have analyzed the sequences of parts of speech for the above mentioned probability, these probability have never been utilized in grammar checking and correcting system (Schabes Col. 3 lines 14-25 & Fig. 1).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Brill to incorporate deriving a grammar being applicable to unknown sentences or unknown phrases as taught by Schabes to allow for the analysis of any input, particularly in any language and being able to not

only translate but interpret the semantic and syntactic structure of discourse, wherein probabilities that check if grammar is correct based on a sequential sentence input (Schabes Col. 3 lines 14-25 & Fig. 1).

Re claims 16-18, Brill fails to teach the method according to claim 1, wherein the weak annotation is one of a set of candidate semantic tags and an inclusion/exclusion list.

Schabes teaches past limitations and an improvement upon them, wherein Schabes teaches that in the past, in order to ascertain proper usage, the grammaticality of a sentence was computed as the probability of this sentence to occur in English. Such statistical approach assigns high probability to grammatically correct sentences, and low probability to ungrammatical sentences. The statistical is obtained by training on a collection of English sentences, or a training corpus. The corpus defines correct usage. As a result, when a sentence is typed in to such a grammar checking system, the probability of the entire sentence correlating with the corpus is computed. It will be appreciated in order to entertain the entire English vocabulary, about 60,000 words, a corpus of at several hundred trillion words must be used. Furthermore, a comparable number of probabilities must be stored on the computer. Thus the task of analyzing entire sentences is both computationally and storage intensive. In order to establish correct usage in the Subject System, it is the probability of a sequence of parts of speech which is derived. For this purpose, one can consider that there are between 100 and 400 possible parts of speech depending how sophisticated the system is to be.

This translates to a several million word training corpus as opposed to several hundred trillion. This type of analysis can be easily performed on standard computing platforms including the ones used for word processing. Thus in the subject system, a sentence is first broken up into parts of speech. For instance, the sentence "I heard this band play" is analyzed as follows: PRONOUN, VERB, DETERMINER, NOUN, VERB. The probability of this part of speech sequence, is determined by comparing the sequence to the corpus. This is also not feasible unless one merely consider the so-called tri-grams. Tri-grams are triple of parts of speech which are adjacent in the input sentence. Analyzing three adjacent parts of speech is usually sufficient to establish correctness; and it the probability of these tri-grams which is utilized to establish that a particular sentence involves correct usage. Thus rather than checking the entire sentence, the probability of three adjacent parts of speech is computed from the training corpus (Schabes Col. 8 lines 13-51).

Further, Schabes teaches that the entries of a dictionary are selected and ranked based on the part of speech assigned to the given word in context. The entries corresponding to the word in context are first selected. The other entries not relevant to the current context are still available at the request of the user. The part of speech of the given word in context is disambiguated with the part of speech tagger described above. By way of illustration, assuming the word "left" in the sentence "He left a minute ago", the part of speech tagger assigns the tag "verb past tense" for the word "left" in that sentence. For this case, the Subject System selects the entries for the verb "leave" corresponding to the usage of "left" in that context and then selects the entries for "left"

not used in that context, in particular the ones for "left" as an adjective, as an adverb and as a noun (Schabes Col. 24 lines 45-60).

Schabes teaches a sentence that is annotated with a tag, such as a part-of-speech tag (Schabes Col. 23 lines 18-25).

Schabes also teaches a listing of words that DO and DO NOT correspond to a proper relationship, wherein Schabes teaches the break down of the proper output of tags, where order to display the entries from the dictionary that correspond to the context, all the entries in a dictionary 970 that correspond to a root found in the set 37 of pairs of roots and parts-of-speech that correspond to the context 950 are displayed at 980. In the above example, all entries for the verb "leave" will be displayed as entries relevant to the context. In order to display the entries from the dictionary that do not correspond to the context, all the entries in the dictionary 970 that correspond to a root found in the set of pairs of roots and parts-of-speech that do not correspond to the context 960 are displayed at 990. In the above example, all entries for the word "left" as an adjective, as an adverb and as a singular noun are displayed as entries not relevant to the context (SchabesCol. 26 lines 12-21).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Brill to incorporate the method according to claim 1, wherein the weak annotation is one of a set of candidate semantic tags and an inclusion/exclusion list as taught by Schabes to allow for a set of sentences in which the words of each sentence are annotated with their part-of-speech tags (Schabes Col. 23

lines 18-25) that improve a NLU expectation-maximization algorithm to retain or omit sets of tags such as elements 980 and 990 (SchabesCol. 26 lines 12-21 & Fig. 14).

Conclusion

4. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Michael C. Colucci whose telephone number is (571)-270-1847. The examiner can normally be reached on 9:30 am - 6:00 pm, Monday-Friday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (571)-272-7602. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Michael C Colucci/
Examiner, Art Unit 2626
Patent Examiner
AU 2626
(571)-270-1847
Michael.Colucci@uspto.gov

/Richemond Dorvil/
Supervisory Patent Examiner, Art Unit 2626